A STATISTICAL METHOD FOR THE TREATMENT OF SCHOOL-SURVEY DATA

L. L. THURSTONE

Bureau of Applied Psychology, Carnegie Institute of Technology,
Pittsburgh, Pennsylvania

The object of the following pages is to present a statistical method which shall enable the city school superintendent to apply educational test data directly in the solution of those administrative problems which are most intimately concerned with the educational efficiency of the system.

If we consider the administration of a city school system we may discriminate between that administrative work which is concerned with the teaching and learning processes as such on the one hand, and all the other administrative work which is concerned with related but non-educational matters such as the size and location of windows, janitors' salaries, the per capita cost of teaching sewing, school lunchrooms, school taxes, and the like. Thousands of these problems are of administrative importance, but the *raison d'être* of the whole school is the teaching and learning processes, and these should be the central topic of school surveys and school reports.

This strictly educational aspect of school administration may be divided into three separate problems: (1) What is to be taught? This is of course the question of deciding upon the content of the curriculum. (2) How should it be taught? This is the question of deciding upon teaching methods. (3) How efficiently is it actually being taught in this particular city at the present time? This category may be called "educational accounting" to distinguish it from other forms of accounting in which we keep tab on the efficiency of the financial part of the administration.

Closely related to this category of "educational accounting" is that of "educational research." If the relative efficiency of two teaching methods were experimentally investigated, the results

would be generally applicable, and to that extent they would constitute a contribution to educational science. But if we investigate by periodical tests the relative efficiency of a particular school system, a particular school, a particular teacher, or the individual pupil, such work is of course not generally applicable. It does not constitute scientific research. It is simply determining the efficiency with which some particular work is carried out. Such work may be more properly called educational accounting.

In educational accounting there are four separate problems. These involve four types of comparison:

1. The relative rank of the school system as a whole when compared with the teaching efficiency of a large group of school systems, measured by means of educational tests.

2. The relative rank of a school within the city when compared with the teaching efficiency of all the other schools in the same city, the measurements being made by means of educational tests.

3. The relative rank of each teacher in the city when compared with the teaching efficiency of all the other teachers of comparable class in the same city.

4. The relative rank of each pupil in the city when compared with the school and test performance of all other pupils of comparable grade in the same city.

An unfavorable report in the first comparison gives warning to the superintendent that he is not succeeding in maintaining the average scholarship standards of other comparable school systems.

An unfavorable report in the second comparison gives warning to the principal concerned that he is not maintaining the standards of other schools of comparable class.

An unfavorable report in the third comparison serves to warn the individual teacher that her teaching efficiency in the subject concerned is below the average of other teachers in the same city.

The reports concerning individual pupils would be of great service as auxiliary data for determining promotion and the skipping of grades.

Let us consider first the comparison of the school system as a whole with other comparable systems. The data for the accompanying figures, which are only illustrative of the method, were

selected from the Des Moines Annual Report, 1915. In this report (p. 17) we find a table of city average scores in fifth-grade spelling tests, for a number of cities. In Fig. 1 these scores are arranged on the axis of abscissae (horizontal axis). The percentile ranks of the cities are arranged on the axis of ordinates (vertical axis). The use of the chart may be best illustrated by an example. Suppose that the average score in a spelling test given to all the

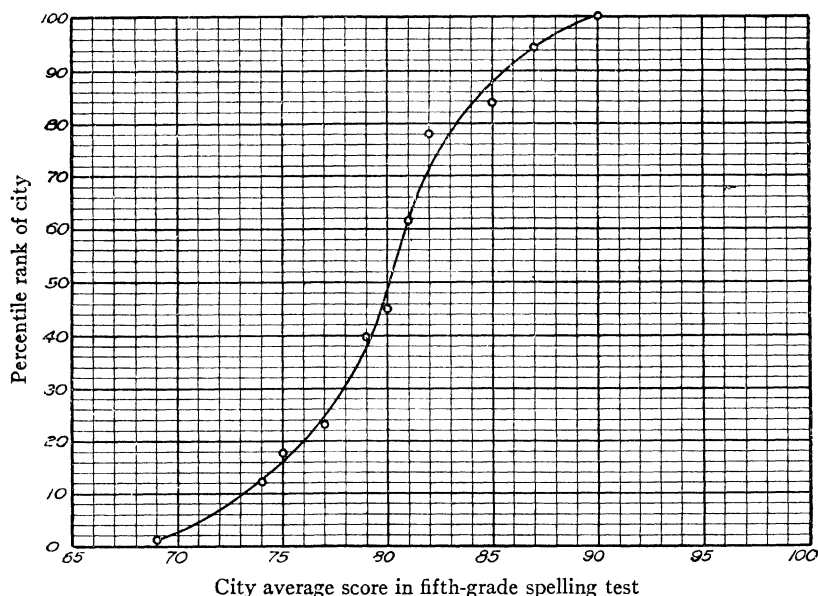City average score in fifth-grade spelling test

FIG. 1

fifth-grade pupils in a city is 85. Then that city will have a percentile rank of 88 per cent as read from the chart. By this rank of 88 per cent is meant that among one hundred cities, the city in question ranks eighty-eighth from the bottom. There are, then, 12 cities among 100 whose fifth-grade pupils excel in spelling, and 88 cities whose fifth-grade pupils do worse in the same spelling test. If the city average in the fifth-grade spelling test be 80, the city would rank (according to the chart) about 50 per cent, indicating that there are as many cities whose fifth-grade pupils do better as there are cities whose fifth-grade pupils do worse in spelling.

Obviously one city may have a high percentile rank in fifth-grade spelling and at the same time have a low percentile rank in fifth-grade arithmetic. Similarly, the same city may have a high percentile rank in one grade and a low rank in another grade. In order to ascertain the relative rank of a city school system, it would be necessary to give to all the pupils a representative set of tests, and to rate the city system on the basis of an average in which the tests for the various school subjects are pooled. In this manner a percentile chart may be drawn up indicating the relative percentile

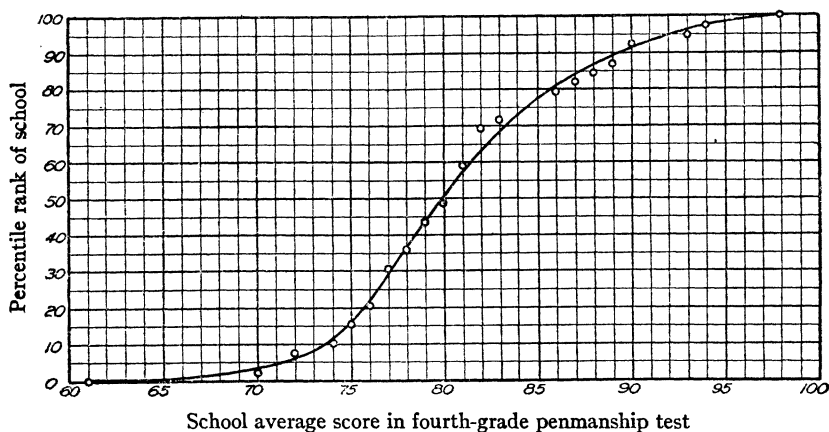School average score in fourth-grade penmanship test

FIG. 2

rank of the city school system as determined by the averages for all the school subjects in all the grades.

In Fig. 2 we have a percentile chart of the various schools in the same city. The data for this table were taken from the Des Moines Annual Report, 1915, p. 83. This particular chart shows the relative percentile rank of the schools within the city system according to scores in a fourth-grade penmanship test. Thus, if a certain school in Des Moines obtains a fourth-grade penmanship average score of 85, the percentile rank of that particular school with reference to fourth-grade penmanship would be about 77 per cent, indicating that 23 per cent of all the schools in Des Moines have better fourth-grade penmanship scores, and that 77 per cent of all the schools in Des Moines have poorer fourth-

grade penmanship scores. Thus the efficiency of any particular school may be measured with reference to all the schools of the city. Obviously the method can be applied to any particular subject, or to any particular grade. The median rank in fourth-grade penmanship among all the schools is, according to the chart, about 80. Any school having a fourth-grade penmanship average of 80 is excelled by as many schools in Des Moines as it exceeds.
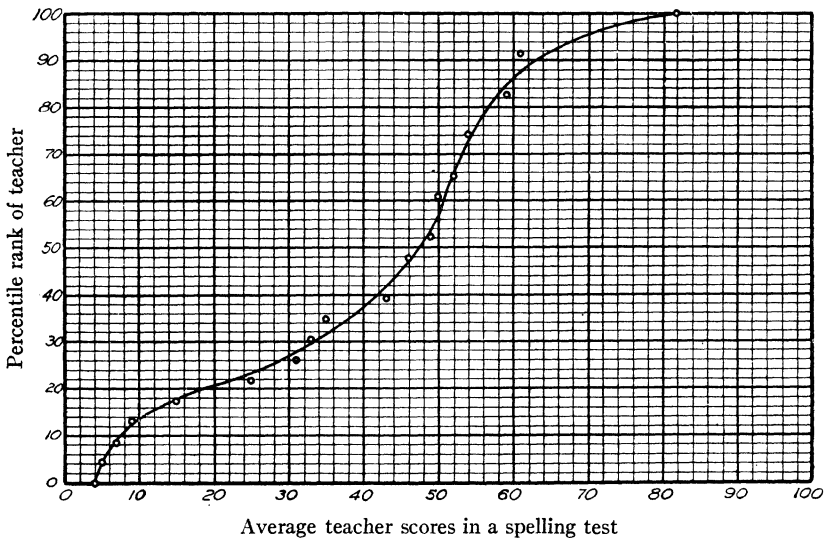

Average teacher scores in a spelling test

FIG. 3

In Fig. 3 the same kind of chart is plotted for the teachers according to their efficiency in the teaching of spelling. The data are taken from the same report (p. 6). Thus a teacher whose class obtains an average spelling-test score of 30 has a percentile rank of about 27 per cent. This particular teacher is exceeded by 73 per cent of the teachers of Des Moines, while she exceeds 27 per cent of the teachers in the same city in the efficiency of teaching spelling. The average for the whole group is always the score which corresponds to the 50 per cent rank which is in this case a spelling score of 47. Thus any teacher whose class obtains a spelling score of 47 exceeds as many teachers in Des Moines as there are teachers who excel her in the ability to teach spelling.

Obviously the results of such tests as these should be published either in chart form or in the form of tables, so that all the teachers throughout the city may compare the scores obtained by their classes with charts or tables and thus determine just where they stand in teaching efficiency as compared with all the other teachers in the same city. The effect on teaching efficiency would certainly be advantageous.
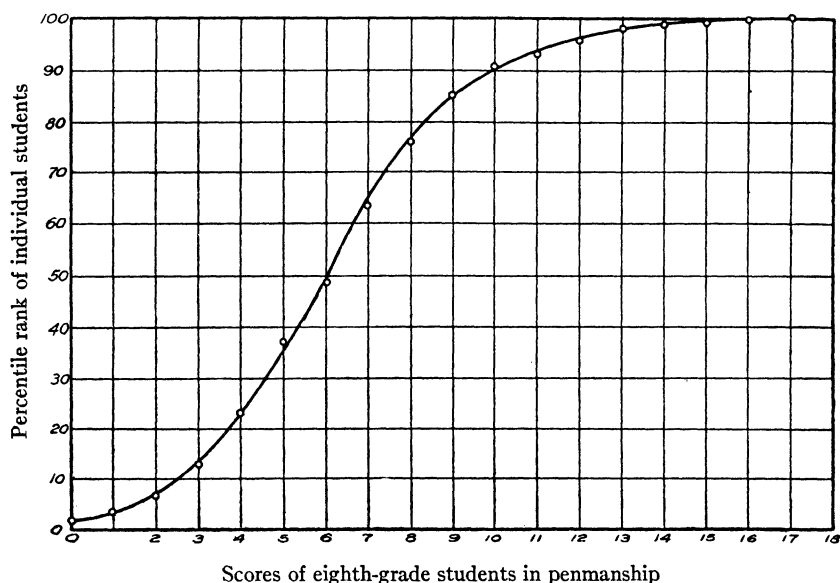

Scores of eighth-grade students in penmanship

FIG. 4

In Fig. 4 we have the norms of performance of eighth-grade pupils in the city of Des Moines in a penmanship test. (See Des Moines Report, 1915, p. 64.) By means of this chart the relative rank of any eighth-grade pupil with respect to all other eighth-grade pupils in Des Moines can be ascertained at a glance. Thus an eighth-grade pupil who scores 8 in this particular penmanship test would rank about 76 per cent, thus indicating that 24 per cent of all the eighth-grade pupils in Des Moines do better in penmanship, while 76 per cent of the eighth-grade pupils in the same city do worse in this subject.

Obviously these records would be of great service in determining whether several teachers, or several schools, or several cities are using comparable standards in their marking systems. The method could also be used to very great advantage as auxiliary data for determining whether a certain pupil is able to skip a grade. When, for example, a certain pupil who is now in the sixth grade obtains a percentile rank of fifty per cent in all his subjects as rated by the seventh-grade pupils, then the rate of promotion of that pupil should be accelerated so as to place him in that grade where, according to his educational test scores, he most properly belongs. The matter of promotion and acceleration would then be less subject to the hazards of personal opinions of teachers. That the educational tests should not constitute the sole means of determining promotion is, of course, obvious.

In stating the problem in this way we are, of course, assuming that we are able to devise fair tests which, when given to school children, will reveal the efficiency of the teaching. On this question there is room for difference of opinion. We may mention several factors which would naturally be thought of as tending to vitiate conclusions drawn from formal educational tests:

a) That the formal educational test does not measure the inspirational and personal influence of a teacher, and that the tests are necessarily limited to the measurement of the efficiency with which the factual material is pounded into the children.

b) That the standard with which the teaching efficiency is compared is the average of the group and not the ideal one toward which the teachers should strive.

c) That two teachers who are equally efficient may have to work with classes which are not of comparable mental caliber, and that the educational tests would give an unfair showing for the teachers who happen to have the less talented children.

The objection to educational tests on the basis that they do not measure the inspirational and personal influence of the teacher on the children, but only the efficiency with which the fundamental facts are learned, may be fairly answered by admitting the limitations of the educational tests. They do not, of course, attempt to measure anything beyond the efficiency with which the factual

material of the instruction has been mastered. But at the same time it must be admitted that the mastery of the factual material of school instruction, especially in the public schools, is one of the most important, if not the most important, object of the teaching work. Any efforts to quantify the efficiency with which this part of the teaching work is carried out must to that extent be laudable. In addition to this consideration we have the fact that a teacher who is efficient in teaching the factual material is in general the one who has the best personal influence on the children. And vice versa, a teacher who is inefficient in giving the factual material is not likely to be inspirationally and personally superior. The exceptions would probably prove the rule.

One might object, at first sight, that this method of measuring teaching efficiency tacitly sets as its standard the average teaching efficiency of the group of teachers considered rather than a high ideal standard. Not exactly. If the teachers whose classes have average scores considerably below the average of the whole group are notified and subsequently improve their teaching efficiency, the obvious result will be that the average for the whole group will also rise. This sets a higher standard of teaching efficiency for the next test and so on. Hence while the basis of comparison is always the average teaching efficiency for the whole group, that average is not a stationary one, since the teachers represented by the scores below the average either improve or gradually become eliminated, thus continually raising the average score for the whole group. Furthermore, such a standard would be more effective in eliminating inefficient teaching than setting up a high ideal standard of performance to be striven for. The average for the whole group of teachers constitutes a standard for which every teacher in the system is to some extent responsible. Its fluctuation will depend upon their own efforts, and their retention and promotion will depend on their relation to this actual and real standard.

A legitimate objection to this method of classifying teachers is that the score obtained by a teacher depends, not only on the teaching ability of the teacher, but also on the mental caliber of the pupils of the class. Two teachers, one of whom has a class of rather dull pupils with unfavorable home influence while the

second has a class of bright children from homes of favorable influence, can only be rated as having equal teaching ability if the second teacher obtains a considerably higher class average in the educational tests than the first. This compensation for the mental caliber of the class can be determined empirically by means of general-ability tests which do not depend for their successful solution on the purely informative side of teaching, but rather on the general intelligence of the child. If we have available the records of general-ability tests together with the educational tests, which latter apply to specific subjects, it will be possible to find the proper correction coefficient to be applied to the percentile rank of a teacher. If the class is below the average in general ability the average score for the teacher should be raised. If the class is above the average in general intelligence, the class average score for the teacher should be reduced. The precise mode of applying this principle must await the accumulation of records of general ability and of educational tests before the correction can be more precisely stated. The correction will be small in cities where the children from the different sections of the town are fairly homogeneous. It will be larger where the children from different sections of the town differ considerably in general ability.

Even at the present time, lacking complete data, we could arrive at an empirical correction formula from the variations in retardation conditions throughout the town. To begin with, the method of percentile ranks for teachers could profitably be applied in cities where the bright and dull pupils are fairly distributed among the several schools of the city.

It is, of course, obvious that considerable research work remains to be done before the educational tests will be sufficiently standardized to be considered satisfactory for extensive work. The point which I have tried to make is that in the educational tests and in the percentile arrangement of teachers according to the average scores obtained by their classes, we have available a method which, even in its present form, would be of great service in the inspectorial work of our public schools.